DOCUMENT RESUME

ED 170 371 TM 008 961

AUTHOR Gilmer, Jerry S.
TITLE The Effects of Immediate Feedback Versus Traditional No-Feedback in a Testing Situation.
PUB DATE 12 Apr 79
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Answer Keys; Answer Sheets; *College Students; Complexity Level; *Feedback; Higher Education; *Low Achievement Factors; Multiple Choice Tests; Student Ability; Student Testing; Test Items; *Test Results
IDENTIFIERS *Test Anxiety

ABSTRACT

Sixty college students from classes in educational measurement were divided into two groups. Each group was administered the same criterion test except that one group received feedback after every item and the other received no feedback. The students were also divided into three ability levels. Each test item was classified two ways; by item difficulty, and taxonomy (cognitive skill) classification. Scores for feedback students were significantly lower than the scores for no-feedback students. There were no interactions involving the feedback variables. These results are consistent with some published research and inconsistent with others. Although anxiety was not a variable in this study, examination of other research tends to support a relationship between feedback effects and anxiety. The combination of higher anxiety and immediate feedback appears to be associated with lower performance, especially in low ability students. (Author/MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED170371

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Effects of Immediate Feedback Versus Traditional No-Feedback in a Testing Situation

Jerry S. Gilmer
University of Iowa

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jerry S. Gilmer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper Presented at

Annual Meeting of

the American Educational Research Association

San Francisco, California

April 12, 1979
Session 33.01

TM008 961

The Effects of Immediate Feedback Versus Traditional No-Feedback in a Testing Situation

Abstract

Sixty students from classes in educational measurement were divided into two groups. Each group was administered the same criterion test except that one group received feedback after every item and the other received no feedback. The students were also divided into three ability levels. Each item was classified two ways; by its item difficulty and taxonomy (cognitive skill) classification. Scores for feedback students were significantly lower than scores for no-feedback students. There were no interactions involving the feedback variable. These results are consistent with some published research and inconsistent with others. Although anxiety was not a variable in this study examination of other research tends to support a relationship between feedback effects and anxiety. The combination of higher anxiety and immediate feedback appears to be associated with lower performance.

The Effects of Immediate Feedback Versus Traditional No-Feedback in a Testing Situation

The question as to the facilitative or debilitative effect of immediate feedback in learning and testing situations remains unanswered. Indeed, support for almost any viewpoint can be well documented using selected literature. No single device for giving Knowledge of Results, or KR, as feedback has been called in recent literature, has been shown to consistently produce facilitative effects. Pressey (1950) developed a punchboard for giving KR and reports that KR facilitates learning. Studies by Angell (1949) and Beeson (1973), also using a type of punchboard, indicate similar results. Beeson utilized KR during testing. Bierbaum (1965) and Spencer and Barker (1969), again using a type of punchboard during testing, report that the KR groups performed significantly lower than the no-KR groups. Montor (1970) and Heald (1970) used the Trainer-Tester response card (the subjects erase a black rectangle on the card) and found no significant difference between right/wrong feedback and no-feedback. Strang and Rust used a similar device for giving feedback and found that the immediate feedback group performed significantly lower than the no-feedback group.

Studies involving programmed instruction, which is available in the form of programmed texts or Computer Assisted Instruction, report conflicting results. Grundin (1969) examined 13 studies involving programmed instruction and found that there were no significant KR effects in any of them. Anderson, Kulhavy, and Andre (1971) used CAI in two experiments involving several feedback conditions. The students who

received right/wrong feedback did significantly better on the paper-and-pencil criterion measure than the no-feedback group in both experiments. In a subsequent similar study by Anderson, Kulhavy, and Andre (1972) the difference between the two groups was in the same direction as the previous study, but was not found to be significant. Other studies involving the use of computer materials (Betz, 1975; Prestwood & Weiss, 1978) indicate no significant performance differences between the KR and no-KR groups, while Betz & Weiss (1976a) report that the KR group performed significantly better. Two other studies involving the computer (Gilman, 1969; Tait, Hartley, & Anderson, 1973) indicate significant positive results for KR when KR is modified from simple right/wrong types of feedback. Even if studies involving the computer were to show consistent positive effects of KR not every institution would be able to afford the terminals. Therefore, further research is necessary regarding the effectiveness of less costly methods of giving immediate feedback. This paper reports on a study which used the Trainer-Tester answer form for giving immediate feedback. This form can be used with a test that contains up to 40 four-alternative multiple-choice items. The students use the eraser end of their pencils to remove a black rectangle which covers a three digit code. The student knows if the item is correct or incorrect by comparing the code to a three-digit number next to the item in the test paper. Each of the 160 rectangles on the form contains a unique code so different students can be given different feedback depending on the answer they choose. Based on the studies of Montor (1970), Heald (1970), and Strang and Rust (1973), where the Trainer-

Tester or a similar device was used, it is expected that the KR group will not perform significantly better than the no-KR group.

Some researchers have investigated KR by levels of ability. Anderson, Kulhavy, and Andre (1971) did not find a significant interaction between KR and verbal ability. In an experiment involving a conventional test and an adaptive test (Betz & Weiss, 1976a) the KR group taking the adaptive test had significantly higher scores only for the high ability group. There was not a significant KR by ability interaction on the conventional test. In the present study subjects were divided into three ability levels to examine the KR by ability interaction.

Two variables regarding the items themselves were also considered in this study; the difficulty of the items and the taxonomy classification of the items (Bloom, 1956). Prestwood and Weiss (1978) reported a KR by difficulty interaction. The KR group scored highest on the most difficult test and lowest on the least difficult test, while the no-KR group scored highest on the least difficult test and lowest on the most difficult test. Although taxonomy classification was not treated as a variable by Anderson, Kulhavy, and Andre (1972) or by Pressey (1950) these researchers noted that their criterion measures consisted primarily of items requiring application or recall skills. Angell (1949) classified his items as knowledge, application of non-quantitative skills, or application of quantitative skills, and reported no significant difference.

Briefly, then, this study examines the main effect of providing immediate feedback with the Trainer-Tester answer form, and the

interaction of immediate feedback with ability, item difficulty and item taxonomy classification.

Method

Subjects

The subjects were 60 college students from two classes in Educational Measurement, one summer and one fall. It was felt that the data for the two classes could be combined because there was virtually no difference between the two groups in their performance on the 36-item criterion measure (mean for summer, 22.91; mean for fall, 22.24).

Materials

The Trainer-Tester answer form was used to provide immediate feedback. The same 36-item multiple-choice test was given to students in each of the two conditions: feedback and no-feedback. The only difference between the test forms that each group received was that the test booklet for the no-feedback group did not contain the three-digit code next to each item.

The Variables

Feedback. There were two feedback conditions, immediate feedback and no-feedback. The no-feedback students circled their selected answers in the test booklet. The immediate feedback students were instructed to keep searching for the correct answer if they got an item incorrect.

Ability. The students were classified into one of three ability groups based on the results of the first test they had taken in the course. The low, middle, and high ability groups contained 22, 20, and 18 subjects, respectively. Half of the subjects in each ability group were randomly assigned to each of the feedback conditions.

Difficulty and Taxonomy. The 36 items were obtained from an item bank that listed an item difficulty value for each item. An item was classified as belonging to difficulty level one, two, or three if its difficulty value was greater than .7, between .4 and .7, or less than .4, respectively. The items were also classified into one of three levels based on the primary cognitive skill required to answer the question. Items were classified as belonging to taxonomy level one if they required knowledge or comprehension, level two if they required application or computation, and level three if they required analysis or evaluation. The items were chosen so that each difficulty level contained four items from each taxonomy level.

The test was given the day before the students took their regular classroom test over the same material.

Dependent Variable. The dependent variable was an arcsine transformation of the proportion of items each student answered correctly in each of the nine difficulty by taxonomy cells. Each subject, therefore had nine scores. The following formula was used (Freeman & Tukey, 1950); new score $y = \text{arcsine}\,[x/(n+1)]^{1/2} + \text{arcsine}\,[(x+1)/(n+1)]^{1/2}$, where x is the number of items correct, and n is the number of items, four in this case. This was a 2 x 3 x 3 x 3 design with feedback and ability between-subjects variables and difficulty and taxonomy within-subjects variables.

Results

The results of a four-way analysis of variance failed to show either a significant interaction involving ability or a significant main effect

for ability. Therefore, ability was not considered in a subsequent three-way analysis of variance. The results of the three-way analysis are summarized in Table 1. There were significant main effects for each of the three variables and no significant interactions involving feedback.

Insert Table 1 about here

The only significant interaction was between difficulty and taxonomy. The overall test means for the feedback and no-feedback groups were 20.2 and 24.2, respectively. The difficulty by taxonomy cell means are reported by feedback group in Table 2.

Insert Table 2 about here

Discussion

The most important result of this study is the poor performance by subjects receiving immediate feedback. These results do not conflict with other studies which used similar forms of immediate feedback (Heald, 1970: Montor, 1970; Strang and Rust, 1973). Some negative comments about the Trainer-Tester answer form made by students who used it suggest that students in the feedback group experienced more anxiety and nervousness than students in the no-feedback group. There is sufficient evidence in the literature to suggest a relationship between anxiety and performance when immediate feedback is given. Bierbaum (1965) reports that the lower scores by the KR group may have been a result of increased anxiety. Strang and Rust report lower scores and more

nervousness for subjects in the KR group. The two studies by Betz and Weiss (1976a, 1976b), when examined together, suggest that anxiety affects low ability subjects more than it affects high ability subjects. Subjects were given either a conventional test or an adaptive test. Significantly higher levels of anxiety were reported by subjects taking the adaptive test. Also, on the adaptive test, there was no significant difference in performance between KR and no-KR groups for low ability subjects. Whereas, on the same test high ability subjects in the KR group had significantly higher scores than the no-KR group. On the conventional test, where anxiety was lower, the KR group had significantly higher scores across both ability levels. In general, immediate feedback, when it is associated with increases in anxiety, appears to be detrimental to students' performance on tests, especially in low ability students. Hansen (1974) reports that high anxiety "appears to interfere with the learner's capacity to utilize the feedback information effectively in performing the task requirement" (p. 253). Bierbaum suggests that anxiety may have increased as a result of the requirement that subjects keep searching for the correct answer when they are informed that their choice was wrong. Neither this study nor that of Anderson, Kulhavy, and Andre (1971) report positive results for the KR condition which required students to keep searching for the correct answer. However, Tait et al. (1973) also required students to keep searching for the correct answer and report a significantly higher mean for the feedback group.

The absence of a significant ability by feedback interaction is consistent with the results of Anderson, Kulhavy, and Andre (1971).

Ability was not an experimental variable in the studies of Montor, and Strang and Rust.

The absence of a difficulty by feedback interaction is in conflict with the results of Prestwood and Weiss (1978). Close examination of Table 2 reveals that, as expected the low difficulty items were easier than the middle difficulty items, which were, in turn, easier than the high difficulty items. This pattern was evident across both feedback conditions. A similar pattern existed in the three taxonomy levels for difficulty one and two. In difficulty level three there was a slight reversal between application or computation items and analysis or evaluation items. However, only three of nine possible differences (three for each difficulty level) were significant: knowledge or comprehension items vs. analysis or evaluation items in difficulty one, and knowledge or comprehension items vs. analysis or evaluation items and application or computation items vs. analysis or evaluation items in difficulty three. This lends some support to the contention of Crawford (1968) that there is not necessarily a direct relationship between item difficulty and complexity.

A case can be made, therefore, with a certain amount of confidence, that there is a relationship between immediate feedback and anxiety, probably more of a relationship than exists between immediate feedback and other variables studied. Further research is required to determine if, indeed, the combination of immediate feedback and higher anxiety is detrimental to performance. Research is required in the area of development of a simple, inexpensive device for giving immediate feedback which does not at the same time increase anxiety.

References

Anderson, R. C., Kulhavy, R. W., & Andre, T. Feedback procedures in programmed instruction. Journal of Educational Psychology, 1971, 62, 148-156.

Anderson, R. C., Kulhavy, R. W., & Andre, T. Conditions under which feedback facilitates learning from a programmed lesson. Journal of Educational Psychology, 1972, 63, 186-188.

Angell, G. W. The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry. Journal of Educational Research, 1949, 42, 391-394.

Beeson, R. O. Immediate knowledge of results and test performance. Journal of Educational Research, 1973, 66, 224-226.

Betz, N. E. Prospectus: New types of information and psychological implications. In D. J. Weiss (Ed.) Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A018675)

Betz, N. E., & Weiss, D. J. Effects of immediate knowledge of results and adaptive testing on ability test performance (Research Report 76-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976. (NTIS No. AD A027147) (a)

Betz, N. E., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive ability testing (Research Report 76-4). Minneapolis: University of Minnesota, Department of

Psychology, Psychometric Methods Program, 1976. (NTIS No. AD AC27170) (b)

Bierbaum, W. B. Immediate knowledge of performance on multiple-choice tests. Journal of Programmed Instruction, 1965, 3, 19-23.

Bloom, B. S. (Ed.) Taxonomy of educational objectives, handbook 1: Cognitive domain. New York: McKay, 1956.

Crawford, W. R. Item difficulty as related to the complexity of intellectual processes. Journal of Educational Measurement, 1968, 5, 103-107.

Freeman, M. F., & Tukey, J. W. Transformations related to the angular and the square root. Annuals of Mathematical Statistics, 1950, 21, 607-611.

Gilman, D. A. Comparisons of several feedback methods for correcting errors by computer-assisted instruction. Journal of Educational Psychology, 1969, 60, 503-508.

Grundin, H. U. Response mode and information about correct answers in programmed instruction: A discussion of experimental evidence and educational decisions. In A. P. Mann & C. P. Brunstrom (Eds.), Aspects of Educational Technology (Vol. 3). London: Sir Isaac Pitman and Sons Limited, 1969.

Hansen, J. B. Effects of feedback, learner control, and cognitive abilities on state anxiety and instruction in a computer-assisted instruction task. Journal of Educational Psychology, 1974, 66, 247-254.

Heald, H. M. The effects of immediate knowledge of results and correlation of errors and test anxiety upon test performance. (Doctoral dissertation, University of Nebraska, 1970). Dissertation Abstracts International, 1970, 31, 1621A-1622A. (University Microfilms No. 70-17, 724)

Montor, K. Effect of using a self scoring answer sheet on knowledge retention. Journal of Educational Research, 1970, 63, 435-437.

Pressey, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. Journal of Psychology, 1950, 29, 417-447.

Prestwood, J. S., & Weiss, D. J. The effects of knowledge of results and test difficulty on ability test performance and psychological reactions to testing (Research Report 78-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Spencer, R. E., & Barker, B. An applied test of the effectiveness of an experimental feedback answer sheet (Research Report No. 293). Urbana, IL: University of Illinois, Measurement and Research Division, 1969.

Strang, H. R., & Rust, J. O. The effects of immediate knowledge of results and task definition on multiple-choice answering. The Journal of Experimental Education, 1973, 42, 77-80.

Tait, K., Hartley, J. R., & Anderson, R. C. Feedback procedures in computer-assisted arithmetic instruction. British Journal of Educational Psychology, 1973, 43, 161-171.

Table 1

Analysis of Variance

Source	df	MS	F
Feedback	1	7.25	14.55*
$Error_b$	58	.50	
Difficulty	2	24.49	127.76*
Difficulty x Feedback	2	.07	.38
$Error_w$	116	.19	
Taxonomy	2	8.85	41.94*
Taxonomy x Feedback	2	.15	.71
$Error_w$	116	.21	
Difficulty x Taxonomy	4	1.27	6.18*
Difficulty x Taxonomy x Feedback	4	.08	.43
$Error_w$	232	.20	

*$p < .001$

Table 2

Means and Standard Deviations of Real Scores and Transformed Scores by Feedback Condition and Difficulty by Taxonomy Level

	Feedback		No Feedback	
	Mean	SD	Mean	SD
Low Difficulty				
Knowledge/Comprehension	3.60 (2.41)	.56 (.36)	3.80 (2.55)	.55 (.31)
Application/Computation	3.36 (2.28)	.80 (.46)	3.60 (2.40)	.49 (.34)
Analysis/Evaluation	2.26 (1.70)	.98 (.48)	2.86 (2.02)	1.07 (.55)
Middle Difficulty				
Knowledge/Comprehension	2.43 (1.80)	1.10 (.54)	2.90 (2.04)	1.06 (.55)
Application/Computation	1.80 (1.46)	1.12 (.55)	2.43 (1.78)	1.13 (.59)
Analysis/Evaluation	1.73 (1.43)	1.08 (.53)	2.23 (1.68)	.93 (.47)
High Difficulty				
Knowledge/Comprehension	2.30 (1.69)	1.11 (.57)	2.50 (1.82)	.93 (.47)
Application/Computation	1.30 (1.23)	.79 (.41)	1.86 (1.49)	1.00 (.49)
Analysis/Evaluation	1.43 (1.29)	.85 (.42)	2.00 (1.57)	.90 (.44)

Note. Maximum score = 4.00. The numbers in parentheses are the means and standard deviations of the transformed scores. Each mean is based on 30 observations.